Assessing the impact of discretisation assumptions in a length-structured population growth model

H. Drouineau[a, *], S. Mahévas[a], M. Bertignac[b] and A. Fertin[c]

[a] IFREMER Department EMH, Rue de l'Ile d'Yeu, B.P. 21105, 44311 Nantes Cedex 3, France
[b] IFREMER Centre de Brest, Laboratoire de Biologie Halieutique, STH/LBH, BP 70, F-29280 Plouzané, France
[c] Université de Tours, IRBI UMR CNRS 6035, Parc Grandmont, 37200 Tours, France

*: Corresponding author : Drouineau H., email address : hdrouine@ifremer.fr

Abstract:

Most of the traditional assessment models are age-structured. However, many biological and exploitation processes are more length-dependent than age-dependent, and the required length–age conversion of available data is often not reliable. Consequently, length-structured or age–length structured models have undergone considerable development in recent years. The growth transition matrix used to model the mean growth and growth variability of the population, is of primary importance in a length-structured matrix model. Building this growth transition matrix is not trivial and it is necessary to assess the impact that various assumptions may have to identify robust model structures. In this study, we assess the effects of (1) time and length discretisation, (2) the distribution of individuals within length classes and (3) the statistical distribution used to describe growth variability, by fitting a growth matrix model to individual quasi-continuous simulated growth data. The study quantitatively demonstrates that the choice of the time step and of length class width is the key point when building a length-structured population growth model. The use of a gamma distribution for the growth increments and/or a uniform distribution of individuals within length classes were found to make the model more robust.

Keywords: Length-structured model; Growth; Discrete model; Sensitivity analysis

17

Introduction

19   From deterministic - single cohort analysis (Pope, 1972) to multi cohort analysis, age-structured

20   models (Fournier and Archibald, 1982 ; Virtala *et al.*, 1998 ; Shepherd, 1999) are widely used for

21   fisheries stock assessments. While these models have proved to be useful in estimating population

22   abundance and assessing stocks, several limitations have been pointed out. One limit, often

23   mentioned, is related to the need to convert catch at size into catch at age, which is often carried out

24   using an age-length key. For some species, the difficulties in determining the age at length result in

25   uncertainty in catch-at-age. Moreover, since the ageing techniques are very expensive, few fishes

26   can be aged, further increasing the sampling error and variability in the age-length key. Another

27   source of uncertainty is the size-dependent selectivity function which can not be explicitly taken

28   into account in age-structured model though a age selectivity function may be used as a proxy.

29   Large interindividual and/or interannual variations in size within an age class could thus result in

30   incorrect estimates of fishing mortality by age. Consequently, length-structured models or age-

31   length structured models have undergone considerable development in recent years (Schnute, 1987 ;

32   Schnute *et al.*, 1989a ; Schnute *et al.*, 1989b ; Sullivan *et al.*, 1990 ; Banks *et al.*, 1991 ; De Leo and

33   Gatto, 1995 ; Fournier *et al.*, 1998 ; DeLong *et al.*, 2001 ; Froysa *et al.*, 2002), but are still not often

34   used in routine stock assessments.

35

36   Different mathematical frameworks can be used to model length-structured population dynamics,

37   depending on whether population states are assumed discrete or continuous and whether the

38   projection time is discrete or continuous. Continuous approaches better match biological processes.

39   However, the disadvantage of this approach is the complexity of handling differential equations

40   which do not necessarily have analytical solutions. In addition, available observed data are

41   generally aggregated thereby reducing the advantages of a continuous model. By contrast, discrete

42 approximations of biological processes simplify model implementation and are easier to compute.

43 Discrete approaches require discrete assumptions on continuous processes. Following, discrete

44 growth modelling necessitates decisions on length classes width, time step duration, growth

45 increment model (stochastic or deterministic) and on the distribution of individuals within each

46 length class.

47   Fish growth is a major process of fish biology (Laslett *et al.*, 2002) and is part of the information

48 necessary to estimate stock size and fishing mortality in stock assessments models (Anda-Montañez

49 *et al.*, 1999). The importance of describing jointly mean growth and individual variability in growth

50 for stock assessment process has been widely acknowledged (Laslett *et al.*, 2002 ; Chen *et al.*,

51 2003). Two approaches are classically considered to model growth variability. The most common

52 model is a mean growth by length class given by a growth function, often a von Bertalanffy growth

53 curve, and a stochastic distribution for individual variability around the mean (Sullivan *et al.*, 1990 ;

54 De Leo and Gatto, 1995 ; Cruywagen, 1997 ; Froysa *et al.*, 2002). Another way to describe growth

55 variability is to integrate individual variability in the growth function with stochastic parameters

56 (Sainsbury, 1980 ; Smith *et al.*, 1998 ; Smith and Botsford, 1998 ; Pilling *et al.*, 2002). Several

57 approaches can be used to estimate growth parameters and corresponding variability. Length-at-age

58 can be back-calculated using information contained in fish hard parts, but this approach is often

59 imprecise due to difficulties in detecting growth increments. A second approach is tagging

60 (Shackell *et al.*, 1997 ; Kendall and Nichols, 2002 ; Laslett *et al.*, 2002 ; Eveson *et al.*, 2004 ; de

61 Pontual *et al.*, 2006). This technique is very expensive as many fishes have to be tagged in order to

62 be successful.

63

64   To date, very few studies have looked at the impact of time and length discretisation on the

65 ability of a population model to fit data and to estimate growth parameters. We chose in a first step

66 to work on a theoretical population only submitted to growth (no death and no recruitment) since it

67   is a key process of length-structured model (Chen, 2003). In this study we provide a methodology

68   to assess the impact of several assumptions of time and length discretisation and growth increments

69   distribution in a length-structured population growth model. Data on individual growth trajectories

70   are rarely available. Thus, we produced several synthetic data sets of individual growth trajectories

71   corresponding to several hypotheses on growth increments. Then we fitted length-structured growth

72   population models to each synthetic data set. We provided from this analysis a flexible length-

73   structured population growth model.

74

75   **Data and methods**

1   A length - structured population growth model

76   A matrix formulation of a discrete time length-structured population growth model is presented

77   below. Since we only focus on the growth process, natural and fishing mortality are not described in

78   the model. The length-structured abundance in number at time $t+1$ is given by the following

79   equation:

80     $$N(t+1) = G \cdot N(t)$$

81   where $N(t)$: vector of abundance per size-class at time t and G is the matrix of probabilities of

82   transition between length classes

83   We assumed a constant growth transition matrix over time (i.e., no seasonality has been taken

84   into account). The population is supposed to grow on average according to a von Bertalanffy

85   growth function, parameterised by three parameters L_∞, K and T_0, $l(t) = L_\infty \times (1 - e^{-K(t-T_0)})$. We

86   used an alternative formulation of this equation $l(t+\Delta t) - l(t) = (L_\infty - l(t)) \times (1 - e^{K \times \Delta t})$

87   (Fabens, 1965) where $l(t+\Delta t)-l(t)$ is here the growth increments during a time step Δt of an

88   individual of size $l(t)$ at the beginning of the time step. Growth increments within the population are

89   random variables (Sullivan *et al.*, 1990 ; Cruywagen, 1997 ; Froysa *et al.*, 2002) whose mean is

90   given by the von Bertalanffy equation and whose variance is proportional to the square of its mean

91   (DeLong *et al.*, 2001) (i.e., assuming a constant coefficient of variation of the growth increments):

92
$$\begin{cases} E(X/l)=(L_\infty - l)\cdot(1-e^{K\cdot\Delta t}) \\ \quad V(X/l)=C\cdot E^2(X/l) \end{cases}$$

93   with *X/l* the random variable describing the growth increment for fish of initial size *l* during a

94   time step *Δt* and *C* a constant linking the mean to the variance equals to $CV^2(X/l)$.

95

96   Given the previous assumptions on the growth increments' mean and variance, the coefficients of

97   the growth matrix were calculated according to the following equations:

98   • if fish size is assumed to be at the midpoint size of the departure class *i*, denoted m_i, then

99   the probability g_{ij} of transition to class *j* is given by the following single integral:

100
$$g_{ij}=\int_{(m_j-\Delta l/2)-m_i}^{(m_j+\Delta l/2)-m_i} f(x)dx \quad \text{with } \Delta l \text{ the width of the length classes}$$

101   • if fish size is assumed to be uniformly distributed within the departure class *i*, then the

102   value of the following double integral is the probability g_{ij} of transition from class *i* to class

103   *j*:

104
$$g_{ij}=\int_{m_i-\Delta l/2}^{m_i+\Delta l/2}\left(\int_{(m_j-\Delta l/2)-y}^{(m_j+\Delta l/2)-y} f(x)dx\right)dy \quad \text{with } \Delta l \text{ the width of the length classes}$$

105   with *f* the assumed density function of growth increment.

106   We considered three different continuous distributions (normal, lognormal, gamma) to describe

107   growth increments distribution (Table 1). The Gamma distribution (Table 2) is a two-parameters

108   distribution, entirely defined by its two first moments. It is largely used in modelling thanks to its

109   flexibility in describing many functional forms. More specifically in our context, it can describe

110   growth for small and large fish. The normal and lognormal (Table 2) statistical distributions were

111   also considered.

112   We considered classes of equal width in length and assumed two types of length classes (Table

113   1), which is the most common method to build classes (Sullivan *et al.*, 1990 ; De Leo and Gatto,

114   1995 ; Froysa *et al.*, 2002).

115   Finally, three different time steps were considered in this analysis: monthly, quarterly and annual

116   (Table 1).

117

118   Let us denote *Mod(Lm, Δt, Δl, ODep)* the formulation of the length-structured population growth

119   model assuming a statistical distribution of growth increments *Lm*, a time step *Δt*, length classes of

120   size *Δl*, and an option of distribution of individuals within classes *ODep* (Table 1). A simulation

121   corresponds to a ten-year cohort projection (we denoted T the number of time steps corresponding

122   to a ten-year projection) with an initial length distribution following a normal distribution of mean 5

123   cm and standard deviation of 1.

124

2   Synthetic Data

125   Individual growth data are necessary to assess the impact of both time and length discretisation

126   assumptions in the growth dynamics model. However, no such data were available: consequently

127   we simulated individual growth trajectories. We used a monthly time step to mimic time continuous

128   growth. Shorter time steps were also tested, but the differences in frequencies-at-length were minor.

129   Let us denote *S(K, CV, Ls)* a simulation set of individual growth trajectories performed with, a

130   growth rate *K*, a constant of variance of growth increments *CV* and a statistical distribution of

131   growth increments *Ls*. A set of 10000 individual trajectories over a ten-year period was simulated

132   for each combination of assumptions of growth variability *(K, CV, Ls)* using a monthly time step

133   (Table 1). Two distinct von Bertalanffy growth rates **K** were considered (Table 1), respectively

134   corresponding to a slow growing species and a faster growing species. Three distinct constant of

135   variation were used (Table 1). These simulations aim at mimicking individual growth observations.

136   For each set, the initial length of each individual was generated according to a normal distribution

137   with mean 5cm and standard deviation of 1.

138

139   *3   Analysis*

140   3.1   Experimental design

141   We aimed at assessing the impact of discretisation assumptions on length-structured population

142   growth model ability to fit sets of individual growth trajectories and estimate growth parameters.

143   We used an experimental design (Kleijnen, 1998) that we analysed using standard linear models.

144   Each formulation of the length-structured population growth model *Mod(Lm, Δt, Δl, Odep)* was

145   fitted to each simulated data set of individual growth trajectories *S(**K**, **CV**, **Ls**)* by estimating *Kest*

146   and *CV* which minimize the sum of squared error:

147   $$LS(Kest,CV) = \sum_{i=1}^{I} \sum_{t=1}^{T} \left(f_{est}(t,i) - f_{obs}(t,i) \right)^2$$

148   with *fobs(t,i)* the simulated frequencies of individuals within class *i* at time step *t, fest(t,i)* the

149   frequencies of individuals within class *i* at time step *t* estimated by the matrix model, *I* the number

150   of classes and *T* the number of time steps. This design required 648 simulations (18 sets of

151   individuals growth trajectories times 36 formulations of the length-structured population growth

152   model).

153

154   The minimisation of the function LS(K_{est},CV) was carried out using a genetic algorithm coupled

155   with a quasi-Newton algorithm provided in the Autodif library (Otter Research Ltd. http://otter-

156 rsch.com). The genetic algorithm was used in a first step to provide a starting point not too far from

157 the optimum for the Autodif algorithm.

158   For each experiment, the adequacy of the length-structured population growth model is

159 quantified using two criteria:

160   ●The discrepancy between the K estimated ($Kest$) by the matrix model and the $\boldsymbol{K}$ used to

161   generate the observations data ($\boldsymbol{K}$). It quantifies the goodness of the estimation of K

162   ●The Bayesian Information Criterion (BIC) is computed using the formula

163 $BIC = -2 \cdot LogL + p \cdot \log(n)$ with p=2 the number of estimated parameters (K and CV)

164 $LogL = n \times \log\left(\sqrt{2 \times \pi}\right) + \dfrac{1}{2} \times \left[n \times \left(\log \dfrac{n-1}{LS} - 1 \right) + 1 \right]$ and n the number of observations (number of

165 time steps X number of classes). The BIC is here a measure of model ability to fit the data. The BIC

166 balances the discrepancy between data and predictions (measured by LS) by the number of

167 estimated parameters and the number of observations (which depends on the time step and length

168 classes width in the model).

169

170 3.2 Assessing discretisation assumptions

171   We first used box-and-whisker plots of the estimated growth rate $Kest$ for each modality of the

172 four assumptions in order to visually detect biases. Then, a statistical analysis of the criteria was

173 carried out to quantify the effect of each hypothesis. First, each criterion was analysed separately

174 using linear models step fitted on the transformed criterion, respectively log(|BIC|)=$LBic$ (the

175 absolute value is necessary for the logarithmic transformation because the BIC were negative) and

176 log(|K-Kest|)=LK, where $LBic$ and LK will refer to the two transformed variables. An appropriate

177 model is characterised by a high $LBic$ and a low $|\boldsymbol{K}\text{-}Kest|$. A full experimental design allows

178   inclusion of all the main effect and first-order interaction in the model (Kleijnen, 1998), so that the

179   models may be written under the general form:

$$
\begin{aligned}
LBic \sim{} & \mu + Lm + \Delta t + \Delta l + Odep + K + CV + Ls + Lm:\Delta t + Lm:\Delta l + Lm:Odep + Lm:K + \\
& Lm:CV + Lm:Ls + \Delta t:\Delta l + \Delta t:Odep + \Delta t:K + \Delta t:CV + \Delta t:Ls + \Delta l:Odep + \Delta l:K \quad (1) \\
& + \Delta l:CV + \Delta l:Ls + Odep:K + Odep:CV + Odep:Ls + K:CV + K:Ls + CV:Ls
\end{aligned}
$$

$$
\begin{aligned}
LK \sim{} & \mu + Lm + \Delta t + \Delta l + Odep + K + CV + Ls + Lm:\Delta t + Lm:\Delta l + Lm:Odep + Lm:K + \\
& Lm:CV + Lm:Ls + \Delta t:\Delta l + \Delta t:Odep + \Delta t:K + \Delta t:CV + \Delta t:Ls + \Delta l:Odep + \Delta l:K \quad (2) \\
& + \Delta l:CV + \Delta l:Ls + Odep:K + Odep:CV + Odep:Ls + K:CV + K:Ls + CV:Ls
\end{aligned}
$$

182   with α_i the effect of the assumption i $\beta_{i,j}$ the effect of the interaction between assumption i and

183   assumption j.

184   The coefficients of an effect are constrained to sum to 0.

185   These linear models enable us to quantify the impact of the assumptions of both quality of K

186   estimation and model fit. In addition, we were also interested in selecting the most appropriate

187   model with respect to both objectives. A multiple regression tree (De'Ath, 2002) was built to

188   explain the variations of the two transformed variables (previously normalized and centered) by the

189   different explanatory variables (i.e., different assumptions). The model is (bold terms correspond to

190   assumptions for the quasi-continuous model., other terms correspond to assumptions on the matrix

191   model).:

192

$$
\frac{(Lbic - mean(Lbic))}{sd(Lbic)}, \frac{(LK - mean(LK))}{sd(LK)} \sim \Delta t + \Delta l + Lm + CV + K + ODep + Ls \qquad (3)
$$

194   The tree is grown by repeated binary partitioning of the data, maximizing the Euclidean distance

195   between the two group centroids and minimizes the total Euclidean distance to the centroid within

196   each group. Each partition is characterized by an explanatory variable and the rank of the

197   explanatory variable selections in the tree reflects the magnitude of impact of the variable on both

198   responses.

199

Results

1 Experimental designs analysis

Table 3 and Table 4 summarize the remaining effects of model (1) and (2) after performing

classical Fisher tests and residuals analysis. The R² of the linear models are 0.995 for the *LBIC,* and

0.674 for the *LK*. In those tables, the estimates quantify the strength of the effect of the considered

assumption. The reader should notice that for:

- model (1) the assumptions which have a positive estimate improve the fit to observed
  frequencies-at-length

- model (2) the assumptions which have a negative estimate improve the estimation of the
  growth rate.

1.1   Fitting observed frequencies-at-length (Table 3)

The significant effects of interaction factors are minor compared to the main effects and do not

influence the results. Among the main effects, most factors are significant, but few have a large

estimated value. Length class width and time step duration are the two factors which have the

strongest influence. The best fits were obtained with a monthly time step and 1 cm length classes

width. Results also demonstrated that a large inter-individual variability and a large growth rate are

more easily fitted by the matrix models.

1.2   Ability to estimate the growth rate K (Table 4)

The box-and-whisker plots (Fig. 1) did not show any bias in the estimation of the growth rate,

221 except when assuming a normal distribution for the growth increments. However, they showed that

222 the precision of the estimation varies across the different assumptions. The variability is particularly

223 high for a monthly time step, a normal distribution of growth increments and 5cm classes.

224 The linear models showed that fewer factors have a statistically significant effect on the

225 estimation of the growth rate, however many more factors have a strong impact the estimation. As

226 previously, classes of 1cm greatly improved K estimation. More surprisingly, short time steps do

227 not systematically improve the estimation of the growth rate. In fact, with a monthly (or quarterly)

228 step, individuals have a too small probability to change class (especially when classes are 5 cm

229 wide) so that K tends to be over-estimated to compensate. This highlights the links existing between

230 size classes and time steps (the interaction of the two factors is strongly significant). The choice of a

231 time step should be strongly linked to the choice of the size of classes, both choices being also

232 constrained by the available data.

233 The linear model demonstrates that, in contrast with the *BIC* analysis, a low growth rate and a

234 weak inter-individual variability favour a good estimation of K. Furthermore, if the individuals are

235 uniformly distributed within classes, the estimation is improved, especially with a short time step.

236 Finally, it highlights the fact that gamma and lognormal distributions used in the length-

237 structured population growth model yield very similar results. The gamma distribution tends

238 however to produce better fits than the lognormal even if a lognormal distribution is assumed for *S*.

239 A normal assumption in the length-structured population growth model gives poor estimation of $\boldsymbol{K}$

240 when a gamma or lognormal distribution is assumed for *S*. The gamma distribution for *Mod* give

241 better results than a normal distribution even when a normal distribution is assumed for *S* (this can

242 be checked by summing the estimates of the main effects and of the interaction). Among the three

243 distributions tested, the gamma distribution appears to be the most flexible distribution to represent

244 the variability of the growth increments.

245

246 *2   Multivariate regression tree (Table 5 and Fig. 2)*

247   The first criterion to partition groups of fits is the width of the classes. Small length classes tend

248 to improve the estimation of K and especially the goodness of the fit of frequencies at length. Then

249 the fits are partitioned according to time steps : large time steps degrade the estimation of

250 frequencies at length, and this is amplified with wide length classes. We can notice that monthly

251 and quarterly time steps are opposed to yearly time step The next two partitions concern the

252 assumption on statistical distribution of growth increments: assuming a gamma or lognormal

253 improves the goodness of the fit.

254

255 **Discussion**

256   This study aimed at analysing the impact of various assumptions on a length-structured

257 population growth model, especially time and length discretisation, in order to detect some artefacts

258 caused by inappropriate assumptions in growth model and to provide the most flexible model. This

259 kind of sensitivity analysis is essential in a discrete model and the methods we provide can be

260 applied in many situations. This study is a preliminary analysis to the development of a complete

261 length-structured model for European hake. In this paper, we chose to only focus on growth because

262 it is a critical point in length-structured model of population dynamics. A von Bertalanffy growth

263 curve was used to model mean growth, but more general growth functions can be considered as

264 polynomial, Gompertz, Verhulst, Richards, Schnute. Inter-individual variability was modelled using

265 a statistical distribution around this mean. Theoretically, a growth increment variance could be

266 estimated for each length class but we made this assumption to reduce the number of parameters to

267 estimate. It should be noticed that this relationship between the mean and variance constrains the

268 statistical distributions. A similar analysis has been carried out assuming a growth increment

269 variance proportional to the mean (Sullivan *et al.*, 1990 ; Sullivan, 1992) (not presented here)

270 leading to very similar results. The Beta-Binomial distribution (Ennis and Bi, 1998) was primarily

271 used to model the distribution of growth increments, but rejected because of unsatisfactory results.

272 Individual growth data are necessary to assess the impact of both time and length discretisation

273 assumptions in the length-structured population growth model. Tagging data would have been an

274 interesting data source, but despite a recent tagging survey, the amount of tagging data for European

275 hake, the species we are currently working on, was insufficient. Consequently we decided to

276 simulate individual growth data, trying to mimic an average von Bertalanffy growth and

277 introducing inter-individual variability. It may have been interesting to simulate individual growth

278 trajectories considering a population with stochastic K and L_∞. The results may largely be

279 influenced by the methods used to generate these individual growth trajectories and therefore

280 should not be generalized in the broader context. More generally, the results we obtain are not

281 generic since they depend (1) on the method used to simulate data (2) on the considered model,

282 species and/or fisheries. However, the methodology we provide is generic and is appropriate to

283 assess the impact of any discretisation processes in a discret model.

284 In this analysis, we considered L_∞ was known because it may be approximated by a function of

285 Lmax and it avoids the problem of the correlation between K and L_∞. This assumption has already

286 been used for example in DeLong *et al.* (2001). The sets of von Bertalanffy parameters used in the

287 study are closed to those currently used for stock assessment of European hake (K=0.1) (Piñeiro

288 and Sainza, 2003) and to the recently assumed growth (K=0.2) (de Pontual *et al.*, 2003 ; Kacher and

289 Amara, 2005 ; de Pontual *et al.*, 2006). The growth rate *K* and the coefficient of variation *CV* were

290 estimated by minimising the sum of squared deviations between frequencies estimated by the

291 length-structured population growth model and simulated frequencies-at-length.

292 We chose two criteria to describe the goodness of fit of the length-structured population growth

293 model to frequencies generated by the quasi-continuous model. Those two criteria correspond to

294 two modelling aims. The first objective was to reproduce and/or predict the observed data. In that

295 case, we are less interested in the biological processes under-lying the model than in its ability to

296   reproduce observations. The other objective was to reproduce the biological process driving the

297   model (for example growth). In that case, the quality of the estimates (precision, bias) is much more

298   important. Selecting a model that achieves these two objectives should be of major importance.

299   However, it requires a method combining these two criteria. Indeed, in this study we showed that

300   the best fits (the ones with the lowest *BIC*) are not necessarily the ones which provide the best

301   estimates of unknown parameters. Until now, we have not found any satisfactory quantitative

302   methods to simultaneously assess the effect of the different assumptions on both criteria. A possible

303   solution would be to build a desirability function (Harrington, 1965) combining criteria, but the

304   choice of the function is rather subjective since the user must first define individual desirability

305   functions for each response. Multivariate regression trees provide a convenient graphical tool to

306   investigate the impact of assumptions in the length-structured population growth model on both

307   criteria simultaneously but this is only an explanatory analysis based on hierarchical clustering.

308   The statistical analysis of the two criteria enables identification of robust model structures for the

309   length-structured population growth model. It demonstrated that the choice of the time step and of

310   length class width was the key point when building a length-structured population growth model.

311   These choices are closely linked to available data and are strongly interdependent. A compromise

312   needs to be found to comply with the two modelling objectives of the model. Other assumptions

313   had a significant effect on the estimation of the growth rate. The gamma distribution was the most

314   flexible statistical distribution to describe growth increments variability. Assuming that individuals

315   are uniformly distributed within length classes, rather than centred, also improved significantly the

316   estimation of the growth rate. We only tried two kinds of distribution of individuals within length

317   classes though some other distributions are sometimes used, for example U-shaped distribution

318   which is especially appropriate for wide length classes. It would also be interesting to assess the

319   impact of the number of individual growth trajectories to quantify the number of trajectories

320   required for a reliable growth rate estimation.

321   This study was a preliminary analysis in the development of a length-based population model for

322   the European hake (*Merluccius merluccius*) stock dynamics. The results inform on the selection of

323   appropriate assumptions of time and length discretisation for this model. The European hake

324   dynamic model will have a quarterly time step and 1 cm width classes since these appear to be a

325   relevant compromise and they correspond to the time-scale of available data. A gamma distribution

326   will be used to describe the variability of growth increments and individuals will be uniformly

327   distributed within length classes.

328

329

References

331 Anda-Montañez, A. D., Arreguin-Sanchez, F., and Martinez-Aguilar, S. 1999. Length-based growth

332   estimates for Pacific sardine (*Sardinops sagax*) in the Gulf of California, Mexico. CalCOFI

333   Reports, 40: 179-183.

334 Banks, H. T., Botsford, L. W., Kappel, F., and Wang, C. 1991. Estimation of growth and survival in

335   size-structured cohort data: an application to larval striped bass (*Morone saxatilis*). Journal of

336   Mathematical Biology, 30: 125-150.

337 Chen, Y., Hunter, M., Vadas, R., and Beal, B. 2003. Developing a growth-transition matrix for the

338   stock assessment of the green sea urchin (*Strongylocentrotus droebachiensis*) off Maine. Fishery

339   Bulletin, 101: 737-744.

340   Cruywagen, G. 1997. The use of generalised linear modelling to determine inter-annual and inter-
341   area variation of growth rates : the caperock lobster as example. Fisheries Research, 29: 119-131.

342   De Leo, G. A., and Gatto, M. 1995. A size and age-structured model of the European eel (*Anguilla*
343   *anguilla* L.). Canadian Journal of Fisheries and Aquatic Sciences, 52: 1351-1367.

344   de Pontual, H., Bertignac, M., Battaglia, A., Bavouzet, G., Moguedet, P., and Groison, A. L. 2003.
345   A pilot tagging experiment on European hake (*Merluccius merluccius*): methodology and
346   preliminary results. ICES Journal of Marine Science, 60: 1318-1327.

347   de Pontual, H., Groison, A. L., Pineiro, C., and Bertignac, M. 2006. Evidence of underestimation of
348   European hake growth in the Bay of Biscay, and its relationship with bias in the agreed method
349   of age estimation. ICES Journal of Marine Science, 63: 1674-1681.

350   De'Ath, G. 2002. Multivariate Regression Trees: a new technique for modelling species-
351   environment relationships. Ecology, 83: 1105-1117.

352   DeLong, A. K., Collie, J. S., Meise, C. J., and Powell, J. C. 2001. Estimating growth and mortality
353   of juvenile winter flounder, *Pseudopleuronectes americanus*, with a length-based model.
354   Canadian Journal of Fisheries and Aquatic Sciences, 58: 2233-2246.

355   Ennis, D. M., and Bi, J. 1998. The beta-binomial model: Accounting for intertrial variation in
356   replicated difference and preference tests. Journal of Sensory Analysis, 13: 348-412.

357   Eveson, J. P., Laslett, G. M., and Polacheck, T. 2004. An integrated model for growth incorporating
358   tag-recapture, length frequency, and direct aging data. Canadian Journal of Fisheries and Aquatic
359   Sciences, 61: 292-306.

360   Fabens, A. J. 1965. Properties and fitting of the von Bertalanffy growth curve. Growth, 29:

361   265-289.

362   Fournier, D. A., Hampton, J., and Sibert, J. R. 1998. MULTIFAN-CL: a length-based, age-

363   structured model for fisheries stock-assessment, with application to South Pacific albacore,

364   *Thunnus alalunga*. Canadian Journal of Fisheries and Aquatic Sciences, 55: 2105-2116.

365   Fournier, D., and Archibald, C. P. 1982. A general theory for analysing catch-at-age data. Canadian

366   Journal of Fisheries and Aquatic Sciences, 39: 1195-1207.

367   Froysa, K. G., Bogstad, B., and Skagen, D. W. 2002. Fleksibest- an age-length structured fish stock

368   assessment model. Fisheries Research, 55: 87-101.

369   Harrington, E. C. J. 1965. The desirability function. Industrial Quality Control, 21: 494-498.

370   Kacher, M., and Amara, R. 2005. Distribution and growth of 0-group European hake in the Bay of

371   Biscay and Celtic Sea: a spatial and inter-annual analyses. Fisheries Research, 71: 373-378.

372   Kendall, W. L., and Nichols, J. D. 2002. Estimating state-transition probabilities for unobservable

373   states using capture-recapture/resighting data. Ecology, 83: 3276-3284.

374   Kleijnen, J. P. C. 1998. Experimental design for sensitivity analysis, optimization, and validation of

375   simulations model. *In* Handbook of simulation :principles, methodology, advances, applications,

376   and practice. Banks, J. (Ed.) , New York. 173-224.

377   Laslett, G. M., Eveson, J. P., and Polacheck, T. 2002. A flexible maximum likelihood approach for

378   fitting growth curves to tag-recapture data. Canadian Journal of Fisheries and Aquatic Sciences,

379   59: 976-986.

380   Pilling, G. M., Kirkwood, G. P., and Walker, S. G. 2002. An improved method for estimating

381   individual growth variability in fish, and the correlation between von Bertalanffy growth

382   parameters. Canadian Journal of Fisheries and Aquatic Sciences, 59: 424-432.

383   Piñeiro, C., and Sainza, M. 2003. Age estimation, growth and maturity of the European hake

384   (*Merluccius merluccius* (Linnaeus, 1758)) from Iberian Atlantic waters. ICES Journal of Marine

385   Science, 60: 1086-1102.

386   Pope, J. G. 1972. An investigation of the accuracy of virtual population analysis for the calculation

387   of fishing mortality from catch at age data. International Commission for the Northwest Atlantic

388   Fisheries Research Bulletin, 9: 65-74.

389   Sainsbury, K. J. 1980. Effect of Individual Variability on the von Bertallanffy Growth Equation.

390   Canadian Journal of Fisheries and Aquatic Sciences, 377: 241-247.

391   Schnute, J. 1987. A general fishery model for a size-structured fish population. Canadian Journal of

392   Fisheries and Aquatic Sciences, 44: 924-940.

393   Schnute, J. T., Richards, L. J., and Cass, A. J. 1989a. Fish survival and recruitment: investigations

394   based on a size-structured model. Canadian Journal of Fisheries and Aquatic Sciences, 46:

395   743-769.

396   Schnute, J. T., Richards, L. J., and Cass, A. J. 1989b. Fish growth: investigations based on a size-

397   structured model. Canadian Journal of Fisheries and Aquatic Sciences, 46: 730-742.

398   Shackell, N. L., Stobo, W. T., Franck, K. T., and Brickman, D. 1997. Growth of cod (*Gadus*

399   *morhua*) estimated from mark-recapture programs of the Scotian Shelf and adjacent areas. ICES

400   Journal of Marine Science, 54: 383-398.

401   Shepherd, J. G. 1999. Extended survivors analysis: an improved method for the analysis of catch-at-

402   age data and abundance indices. ICES Journal of Marine Science, 56: 584-591.

403   Smith, B. D., and Botsford, L. W. 1998. Interpretation of growth, mortality, and recruitment

404   patterns in size-at-age, growth increment, and size frequency data. Proceedings of the North

405   Pacific Symposium on Invertebrate Stock Assessment and Management in Can. J. Fish. Aquat.

406   Sci., 125: 125-139.

407   Smith, B. D., Botsford, L. W., and Wing, S. R. 1998. Estimation of growth and mortality

408   parameters from size frequency distributions lacking age patterns: the red sea urchin

409   (*Strongylocentrotus franciscanus*) as an example. Canadian Journal of Fisheries and Aquatic

410   Sciences, 55: 1236-1247.

411   Sullivan, P. J. 1992. A Kalman Filter Approach to Catch-at-Length Analysis. Biometrics, 48:

412   237-257.

413   Sullivan, P. J., Lai, H., and Galluci, V. F. 1990. A catch -at- length analysis that incorporates a

414   stochastic model of growth. Canadian Journal of Fisheries and Aquatic Sciences, 47: 184-198.

415   Virtala, M., Kuikka, S., and Arjas, E. 1998. Stochastic virtual population analysis. ICES Journal of

416   Marine Science, 55: 892-904.

417

418

419   Table 1: List of the abbreviations and their specifications. Bold terms correspond to assumptions
420   on the quasi-continuous model. Other terms correspond to assumptions on the length-structured
421   population growth model.

Abbreviation	Signification	Possible Values
Δt	Time steps of the length-structured population growth model	12: monthly time step 4: quaterly 1: yearly
Δl	Size of the length classes in the length-structured population growth model	1: 1cm 5: 5cm
Lm	Statistical distribution of the growth increments in the length-structured population growth model	0: gamma 1: normal 2: lognormal
$ODep$	Distribution of the individual within length classes in the length-structured population growth model	0: center 1: uniform
C - CV	Constant of variance in the quasi-continuous model - corresponding coefficient of variance	$0.4 - 0.6$ $1 - 1$ $4 - 2$
K	Growth rate in the quasi-continuous model	$0.1, 0.2$
Ls	Statistical distribution of the growth increments in the quasi-continuous model	0: gamma 1: normal 2: lognormal
α	Parameter of the Beta-Binomial	$0.1, 1, 10, 50$
β	Parameter of the Beta-Binomial	$1, 10, 50, 100$
$Kest$	Estimated growth rate when fitting the length-structured population growth model on frequencies at length generated by the quasi-continuous model	
BIC	Bayesian Information Criterion when fitting the length-structured population growth model on frequencies at length generated by the quasi-continuous model	
$Lbic$	Neperian logarithm of the absolute value of BIC	
LK	Neperian logarithm of the absolute value of the difference bewteen $Kest$ and K	

422

423   Table 2: Distribution, mean and variance of the normal (N), lognormal (LN) and Gamma (G)

424   statistical distributions. With $\Gamma(x)=\int\limits_{0}^{\infty} e^{u} u^{x-1} du$ the gamma function.

Distrib	Density	E	V
N	$f(x)=\dfrac{1}{\sigma\times\sqrt{(2\times\pi)}}\times e^{-\frac{1}{2}\times\left(\frac{x-\mu}{\sigma}\right)^2}$	$E(X)=\mu$	$V(X)=\sigma^2$
LN	$f(x)=\dfrac{1}{\sigma\times x\times\sqrt{(2\times\pi)}}\times e^{-\frac{1}{2}\times\left(\frac{\log(x)-\mu}{\sigma}\right)^2}$	$E(X)=e^{\mu+\frac{\sigma^2}{2}}$	$V(X)=(e^{\sigma^2}-1)\times e^{2\mu+\sigma^2}$
G	$f(x)=\dfrac{1}{\beta^{\alpha}} x^{\alpha-1} e^{-\frac{x}{\beta}}$	$E(X)=\dfrac{\alpha}{\beta}$	$V(X)=\dfrac{\alpha}{\beta^2}$

425

426

427   Table 3: Significant effects in the linear model explaining *LBic* of the fits by the different
428 assumptions .
429   *** (respectively **) stands for a p-value<0.01 (respectively <0.05)

	Estimate	P-value
Intercept	8.92	***
Lm=0	0.03	***
Lm=1	-0.05	***
Δt=1	-1.18	***
Δt=4	0.06	**
Δl=1	1.37	***
CV=0.2	-0.21	***
K	0.89	***
Ls=0	0.05	**
Ls=1	-0.06	***
ODep=0	-0.03	***
Δt=1:Δl=1	-0.07	***
Δt=4:Δl=1	0.04	***
Δt=1:K	0.67	***
Δt=1:Lm=0	-0.02	**
Δt=1:Lm=1	0.03	***
Δt=4:Lm=1	0.02	**
Δt=1:Ls=1	0.04	***
Δl=1:K	-0.49	***
Δl=1:Ls=0	-0.04	***
Δl=1:Ls=1	0.06	***
Lm=1:CV=0.2	0.03	***
CV=0.2:K	0.31	**
CV=1:K	0.22	**
CV=0.2:Ls=0	-0.05	***
CV=0.2:Ls=1	0.08	***
K:Ls=1	-0.42	***

430

431   Table 4: Significant effects in the linear model explaining *LK* of the fits by the different
432 assumptions.
433   *** (respectively **) stands for a p-value<0.01 (respectively <0.05)

	Estimate	P-value
Intercept	-4.74	***
Δt=1	-0.49	***
Δl=1	-0.38	***
Lm=0	-0.35	***
Lm=1	0.66	***
ODep=0	0.21	***
CV=0.2	-0.65	***
K	4.93	***
Ls=0	-0.40	***
Ls=1	0.64	***
Δt=1:Δl=1	0.14	***
Δt=1:ODep=0	-0.22	***
Δl=1:ODep=0	-0.13	***
Δl=1:Ls=0	-0.23	***
Δl=1:Ls=1	0.35	***
Lm=0:Ls=0	-0.17	**
Lm=1:Ls=0	0.42	***
Lm=0:Ls=1	0.33	***
Lm=1:Ls=1	-0.68	***
ODep=0:CV=0.2	0.13	***
Lm=1:CV=0.2	-0.27	***
Δt=1:Ls=0	-0.15	**
Δt=1:Ls=1	0.18	**

434

435   Table 5: Information on the nodes of the multivariate regressions trees (Fig. 2). The best fits are
436   the ones with the lowest *LK* and the highest *Lbic*

Node number	Number of adjustements in the node	Mean *Lbic* (centred and reduced)	Mean *LK* (centred and reduced)	Deviance
1	648	0.00	0.00	1294
2	324	0.26	-0.83	363
3	324	-0.26	0.83	72
4	108	-0.17	-1.46	195
5	216	0.48	-0.51	445
6	108	-0.51	0.10	97
7	216	-0.14	1.19	251
8	144	-0.48	1.20	134
9	72	0.53	1.16	26
10	96	-0.96	1.23	41
11	48	0.49	1.15	69

437

438

439   Fig. 1: box-and-whisker plots of the estimated growth rate for the different assumptions of the
440   length-structured population model. Each row of plots corresponds to an assumption of the length-
441   structured population model while each column corresponds to a growth rate value used to generate
442   individual growth trajectories. In each plot, the different modalities of the considered assumption
443   are plotted in abscissa, the estimated growth rate being plotted on the Y-axis. The horizontal lines
444   represent the assumed growth rate for the simulated data.

445   Fig. 2: Multivariate regressions trees of the fits of the length-structured population growth model
446   on simulated frequencies when. n indicates the number of adjustements in the leaves. The
447   parameters are defined in Table 1. The grey numbers indicate the number of the nodes, more
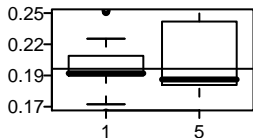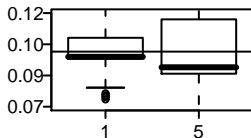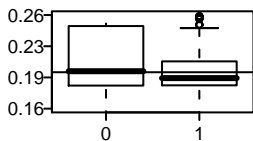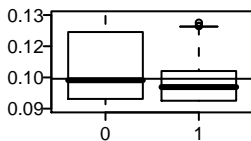448   information on each node can be found in Table 5

Lm

Δt

ΔI

ODep

Kest

K = 0.1   *K* = 0.2